

## Laboratorium statystyczne 2

### Regresja wielokrotna

Dane znajdują się w zbiorze lista 1.xls

#### 1. Dane: *pyłki*.

- Sprawdź<sup>1</sup>, że frakcja zebranego pyłku może być wyrażona w skali oryginalnej.
- Sprawdź poprzez zastosowanie metody strzałki Tukeya<sup>2</sup>, że związek między czasem a frakcją zbieranego pyłku może być liniowy.
- Zbuduj model liniowy, który opisuje związek między frakcją zebranego pyłku, czasem zbierania i czynnikiem opisującym, jaki owad zbierał pyłki.
- Narysuj wykres reszt cząstkowych<sup>3</sup> frakcji względem czasu. Czy postulat liniowości może być podtrzymany?
- Zbadaj czy występuje interakcja między czasem a rodzajem owadów.
- Przeprowadź analizę wariancji dla modeli bez interakcji i z interakcją. Skoryguj, jeśli trzeba, ostateczny model.
- Narysuj, według tego ostatecznego modelu, wykres zależności frakcji od czasu i typu owada, zbierającego pyłek.
- Wykonaj kroki a- c, e, f, g dla frakcji w skali logitowej<sup>4</sup>. Zauważ, że czas powinien być w skali pierwiastkowej.
- Z modelu logitowego odczytaj jak zmieni się<sup>5</sup> frakcja zebranego pyłku przez danego owada, gdy czas zbioru zwiększy się z 16 sek. do 25 sek.
- Ostatecznie, który z modeli dla frakcji wybierasz: w skali oryginalnej, czy logitowej? Uzasadnij odpowiedź, również patrząc na wykresy z punktu g.

#### 2. Dane *nino*

- Sprawdź, że indeks sztormu powinien być wyrażony w skali logarytmicznej
- Znajdź równanie regresji logarytmu indeksu sztormu względem zmiennej indykatorowej *west\_afica* i zmiennych indykatorowych opisujących poziomy występujące w zmiennej *el\_nino* (bez interakcji) Zapamiętaj wartość sumy kwadratów reszt (analiza ANOVA dla regresji) i liczbę stopni swobody. Zapamiętaj wartość sumy kwadratów modelu<sup>6</sup> i liczbę stopni swobody<sup>7</sup>
- Znajdź równanie regresji logarytmu indeksu sztormu względem zmiennej indykatorowej *west\_afica* i zmiennych liczbowych *temperatura* oraz (stwórz taką zmienną) *temperatura*<sup>2</sup>. Zapamiętaj wartość sumy kwadratów reszt (analiza ANOVA dla regresji) i liczbę stopni swobody. Zapamiętaj wartość sumy kwadratów modelu i liczbę stopni swobody
- Wyjaśnij dlaczego te sumy kwadratów i liczba stopni swobody są takie same w b i c.
- Wyjaśnij jak test o tym czy współczynnik przy *temperatura*<sup>2</sup> jest istotny pozwala rozstrzygnąć, który z modeli b czy c użyć. Podaj praktyczną interpretację różnicy między modelami b i c.
- Zbadaj, czy efekt *west\_afica* i czy efekt *el\_nino* jest istotny i czy występują między nimi interakcje.

#### 3. Dane *kentucky*

- Przetestuj model zależności prędkości od roku i warunków na torze wraz z interakcjami. Czy interakcje są istotne? Skomentuj zależność współczynnika przy roku w zależności od warunków na torze.
- Zbuduj alternatywny model zastępując opis *slow, good, fast* przez -1,0,1. Skomentuj zależność współczynnika przy roku w zależności od warunków na torze. Który z tych modeli jest lepszy?
- Narysuj wykres liniowej zależności prędkości od roku w zależności warunków na torze dla obu modeli. Skomentuj,

<sup>1</sup> wykres symetrii, w R: `boxcox(MASS)`,

<sup>2</sup> Metoda strzałki Tukeya.pdf na mojej stronie

<sup>3</sup> W R `prplot(faraway)`

<sup>4</sup>  $\text{logit}(p) = \log(p) - \log(1-p)$

<sup>5</sup> O ile % czy o ile

<sup>6</sup> Suma sum kwadratów przy każdej zmiennej

<sup>7</sup> Suma stopni swobody dla każdej zmiennej